Computer Methods and Programs in Biomedicine 157 (2018) 85–94

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system

Mohammed A. Al-masni[a,1], Mugahed A. Al-antari[a,1], Jeong-Min Park[a], Geon Gi[a], Tae-Yeon Kim[a], Patricio Rivera[a], Edwin Valarezo[a], Mun-Taek Choi[b], Seung-Moo Han[a], Tae-Seong Kim[a,*]

[a] Department of Biomedical Engineering, College of Electronics and Information, Kyung Hee University, Yongin, Republic of Korea
[b] School of Mechanical Engineering, Sungkyunkwan University, Republic of Korea

ARTICLE INFO

Article history:
Received 11 August 2017
Revised 8 December 2017
Accepted 15 January 2018

Keywords:
Breast cancer
Mass detection and classification
Computer Aided Diagnosis
Deep learning
You Only Look Once (YOLO)

ABSTRACT

Background and objective: Automatic detection and classification of the masses in mammograms are still a big challenge and play a crucial role to assist radiologists for accurate diagnosis. In this paper, we propose a novel Computer-Aided Diagnosis (CAD) system based on one of the regional deep learning techniques, a ROI-based Convolutional Neural Network (CNN) which is called You Only Look Once (YOLO). Although most previous studies only deal with classification of masses, our proposed YOLO-based CAD system can handle detection and classification simultaneously in one framework.

Methods: The proposed CAD system contains four main stages: preprocessing of mammograms, feature extraction utilizing deep convolutional networks, mass detection with confidence, and finally mass classification using Fully Connected Neural Networks (FC-NNs). In this study, we utilized original 600 mammograms from Digital Database for Screening Mammography (DDSM) and their augmented mammograms of 2,400 with the information of the masses and their types in training and testing our CAD. The trained YOLO-based CAD system detects the masses and then classifies their types into benign or malignant.

Results: Our results with five-fold cross validation tests show that the proposed CAD system detects the mass location with an overall accuracy of 99.7%. The system also distinguishes between benign and malignant lesions with an overall accuracy of 97%.

Conclusions: Our proposed system even works on some challenging breast cancer cases where the masses exist over the pectoral muscles or dense regions.

1. Introduction

Breast cancer is one of the most leading cancers for women. In 2016, about 246,660 women were diagnosed with breast cancer which is considered as the highest level of 29% among other kinds of cancers [1]. For the expected deaths, breast cancer is the second highest in women which alone accounts 14% against other cancer types [1]. Early detection with correct diagnosis is extremely important to increase the survival rate. In clinical practice, mammography is a widely used diagnostic tool to screen breast cancer. To correctly detect and diagnose breast cancer (i.e., benign or malignant), radiologists face challenges due to the large amount of breast images they have to examine daily and the difficulty of reading them (i.e., detecting the breast masses and correctly diagnosing them). Thus, computer-aided detection and diagnosis (CAD) are essential through which a second opinion can be provided to physicians to aid and support their decisions.

Several studies have been conducted to build CAD systems utilizing conventional recognizers which are attempted to differentiate the breast lesions. In 2016, J. Virmani et al. developed a CAD system to recognize the breast densities [2]. They extracted different statistical texture features from the mass ROIs with different length of Laws' texture energy masks. The dimensionality of these feature vectors was reduced using Principal Component Analysis (PCA). The first four components of the texture features were employed for classification. The results of this CAD system was achieved using Support Vector Machine (SVM) and Probabilistic Neural Network (PNN) classifiers with classification accuracies of 94.4% and 92.5%, respectively. In 2016, C. Muramatsu et al. uti-

* Corresponding author.
E-mail addresses: m.almasani@khu.ac.kr (M.A. Al-masni), en.mualshz@khu.ac.kr (M.A. Al-antari), jmpark@khu.ac.kr (J.-M. Park), geon@khu.ac.kr (G. Gi), kty@khu.ac.kr (T.-Y. Kim), patoalejor@khu.ac.kr (P. Rivera), edgivala@khu.ac.kr (E. Valarezo), mtchoi@skku.edu (M.-T. Choi), smhan@khu.ac.kr (S.-M. Han), tskim@khu.ac.kr (T.-S. Kim).
[1] Equally contributed to this work.

https://doi.org/10.1016/j.cmpb.2018.01.017
0169-2607/

lized texture attributes to distinguish between benign and malignant masses [3]. They proposed an ROI-based feature technique called Radial Local Ternary Patterns (RLTP) which represent the orientation of edge pattern from the center of mass. These RLTP feature sets were compared to ordinary Local Ternary Patterns (LTP), Rotation Invariant Uniform (RIU) LTP, wavelet features, and texture attributes from the Gray Level Co-occurrence Matrix (GLCM). Their CAD system performance of NN overcame Random Forest (RF) and SVM classifiers by 0.9, 0.895, and 0.881 in terms of areas under the receiver operating characteristic curves (AUC), respectively. In 2017, H. Li et al. developed a CAD system based on local contour features to classify benign and malignant masses [4]. They converted the 2D contour of the masses into 1D vector of features. Four different subsections were generated by segmenting the whole 1D signature. New features of Root Mean Square (RMS) slope, describing the contour roughness, were extracted from each subsection besides the fractal dimension and the mean to standard deviation ratio features. Higher classification accuracy of 99.66% was achieved using SVM compared with 99.60% and 92.47% in the case of NN and k-Nearest Neighbors (KNN), respectively. In 2017, S. A. Taghanak et al. proposed a deep auto-encoder network for multi-objective optimization [5]. Their goal was to reduce the dimensionality of features. They extended the conventional auto-encoder to get an optimal solution with more prominent features which in results minimized both mean squared reconstruction and classification errors. Their auto-encoder achieved the classification accuracy of 98.45% for 12 classes. Most of these CAD systems require manual detection of the masses before extracting the features where proper features need to be identified by an expert. This makes any CAD system manual or semi-automatic under clinical settings. Also, these CAD systems could not support detection and classification issues in a single framework. As an alternative to conventional classifiers that utilize hand-crafted features, deep learning techniques can learn prominent features from the entire data [6,7].

For that reason, recently, deep learning is gaining a lot of attention in the field of machine learning. It has been also employed in the field of CAD for breast cancer to overcome some of the limitations of the conventional CAD systems mentioned above. It is considered that deep learning methods can learn a set of high-level attributes and provide a high recognition accuracy instead of using handcrafted features. In 2016, Z. Jiao et al. developed a CAD system based on Convolutional Neural Network (CNN) to classify benign and malignant masses of breast cancer. They utilized the combination of low and high level deep features from two different CNN layers to train their model [6]. Their CAD system succeeded to classify the breast masses with classification accuracy of 96.7%. In 2016, J. Arevalo et al. developed a CNN framework to address the mass lesions of mammograms [7]. The ability of CNN model was investigated against the Histogram of Oriented Gradient (HOG) and Histogram Gradient Divergence (HGD) methods which extracted the features from the histogram. Their CAD system performance achieved AUC of 0.86 compared with 0.796 and 0.793 in the cases of HOG and HGD, respectively. In 2015, N. Dhungel et al. developed an algorithm using a cascade of deep learning and RF to detect the suspicious regions in mammograms [8]. Their algorithm consisted of multi-scale Deep Belief Network (DBN) to select all potential suspicious regions, CNN to keep the correct candidates of those regions, and RF to reduce false positive of the detected regions. Their approach achieved 96% of the true positive cases and 87% of the false positive cases. In 2017, N. Dhungel et al. proposed a total system for detection, segmentation, and classification of the breast masses [9]. They utilized the detected masses from [8] to segment the contours of the actual masses via a deep learning structure followed by Conditional Random Field (CRF). Thereafter, the segmented masses were refined using the Chan–Vese active contour model. Finally, a classical CNN classifier was pre-trained for regressing hand-crafted features and subsequently fine-tuned the pre-trained model. Their system showed an overall segmentation accuracy of 85%. Meanwhile, the performance of their system achieved 91% and 76% in terms of classification accuracy and AUC, respectively. In 2016, T. Kooi et al. employed a deep CNN to classify ROIs for malignant masses [10]. They investigated the power of CNN against the experiences of four radiologists. CNN exhibited its effective ability to recognize the malignant lesions with AUC of 0.87 against 0.84 in the case of radiologists. In 2016, M. Al-antari et al. developed a CAD system utilizing a DBN classifier to distinguish between three different regions of breast cancer (i.e., normal, benign, and malignant), whereas these masses are automatically classified [11]. The capability of DBN was presented against traditional predictors and produced the recognition rate of 92.33%. In 2016, A.-B. Ayelet et al. developed a region-based CNN (R-CNN) method to address the issue of tumor detection and classification [12]. In their work, the mammograms were first preprocessed by removing pectoral muscles and extracting the fibro-glandular. The entire images were divided into multiple overlapped parts. Then, their R-CNN was trained to detect the tumor region and to classify the tumor as benign or malignant. Their results achieved accuracy of 72% and 77% in terms of tumor detection and classification, respectively. In 2017, Y. Qiu et al. built a traditional CAD system to classify the breast masses into benign or malignant [13]. They utilized three convolutional layers to extract the features from 560 resampled ROIs. These features are connected to a multiple layer perceptron classifier with only one hidden layer and one logistic regression layer. Their proposed CAD system produced an overall AUC of 79%. In 2017, G. Carneiro et al. developed an automated deep learning model to examine the two-view of unregistered mammographic images (i.e., CC and MLO) [14]. Both views of each breast image with the segmented maps of their mammogram lesions (i.e., micro-calcifications and masses) were fed in the convolutional network model. Their system achieved 90% and 70% in term of volume under the ROC surface for both semi-automated and fully automated technique, respectively.

In this paper, a novel CAD system is proposed for breast masses detection and classification by employing a novel regional convolutional neural network called You Only Look Once (YOLO) [15]. We augmented the original database of 600 cases by rotating the original mammograms using three different angles for training and testing. YOLO offers a powerful functionality in that it can learn ROIs and their background at the same time. Thus, our proposed CAD system can achieve both detection and classification of breast masses in a single framework. We evaluate the proposed YOLO-based CAD system through using two different datasets (i.e., original and augmented datasets). Our proposed system exhibits an overall accuracy of detection and classification of 99.7% and 97%, respectively.

This paper proceeds as follows. First, we present the overall system with the information of original and augmented databases for training and testing. Second, the details of YOLO-based CAD system for detection and diagnosis of breast cancer masses is explained. Then, we evaluate the performance of our proposed CAD system throughout five-fold cross validation. Finally, we discuss about the results against other CAD works employing classifiers such as DBN and CNN. Finally, conclusion of this work is given.

2. Materials and methods

2.1. Our proposed YOLO-based CAD system

Schematic diagram of the proposed CAD system is demonstrated in Fig. 1. Our proposed YOLO-based CAD system for simultaneous breast masses detection and classification consists of four main stages: mammogram preprocessing, feature extraction

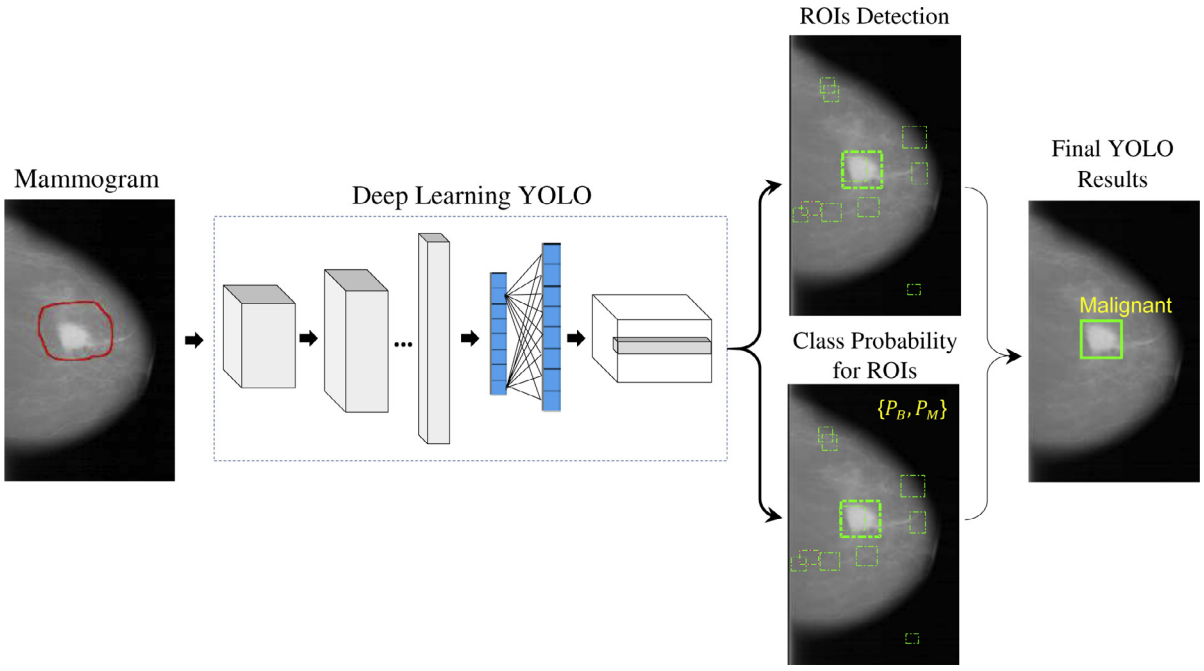

Fig. 1. Scheme of the proposed YOLO-based CAD system.

utilizing multi convolutional deep layers, mass detection with confidence model, and fully connected neural network (FC-NN) for breast mass classification.

2.2. Original database

In this study, we utilized a database of mammograms from Digital Database for Screening Mammography (DDSM) [16] to train and test our YOLO-based CAD system. The DDSM database is created by the University of South Florida and it has been widely utilized in breast research purposes [6,11,17]. It contains 2620 cases which are organized in 43 volumes. Four mammograms are collected for each case with two different views: mediolateral oblique (MLO) and craniocaudal (CC). Each mammogram contains suspicious lesions associated with information of the ground truth. In this work, we have randomly selected a set of 600 mammograms from DDSM database which are equally categorized to benign and malignant cases.

2.3. Augmented database

In fact, deep learning requires large amount of data for proper training. However, small size of medical dataset is one of the most challenging to handle deep learning approaches. Due to this, we used a technique of augmentation to increase the training data. Augmentation is a process that generates new instances from the original data using different transformation methods such as rotation, translation, and scale [6,18,19]. In order to minimize the overfitting problems, that may appear when small size of dataset is utilized via deep learning techniques, we have augmented our dataset three times by rotating the mammograms with angles of 90°, 180°, and 270° as successfully applied in [6,14,18,19]. Thus, a total of 2400 mammograms (i.e., the original mammograms along with their augmented data) are used to train and test the proposed YOLO-based CAD system. The half of mammograms represents the benign and the other half for the malignant. All original and augmented mammograms are randomly mixed together in order to avoid any classification bias of our CAD system.

2.4. Data preprocessing

In this work, mammograms and their ROIs (i.e., masses) must be learned by YOLO. In preparation of input data, we first applied the multi-threshold peripheral equalization technique [20,21] to remove the effect of breast compression that occurred during the examining stage [22]. The peripheral density correction is achieved by the following steps. First, the mask containing breast region is generated using the Otsu thresholding technique. Then, the mask image is multiplied with the blurred image which is produced by applying 2D Gaussian low pass filter to original breast image. Then, the normalized thickness profile (NTP) is derived using different threshold values. These threshold values are computed with respect to the average of blurred image. Finally, the peripheral density correction image is achieved by dividing the original mammogram over the NTP image [23]. This procedure improves the characteristics of the mammograms by eliminating the background and irrelevant data as presented in the previous work [11,23]. In order to achieve a high performance of CAD system, training and testing datasets are normalized in the range of [0, 1] as presented in [6,24]. In DDSM, the mammograms exist with different image sizes [16], hence training and testing datasets are resized to a size of 448×448 as in [15].

2.5. What is YOLO?

You Only Look Once (YOLO) is one of the state-of-the-art deep learning techniques [15]. It is able to detect and classify objects in the entire images at the same time. Unlike previous detection techniques that applied the classifier to multiple regions of the image [8], YOLO utilizes a single convolutional neural network to the whole image. This approach divides the input image into sub-regions and predicts multiple bounding boxes with their class probabilities for each region. Unlike traditional R-CNN that requires many networks for all the extracted regions, YOLO utilizes the entire mammograms so that the contextual information of the predictors and their aspect are completely encoded with a single network in both training and testing time [15]. YOLO has several advantages over other detection systems. This is due to that YOLO

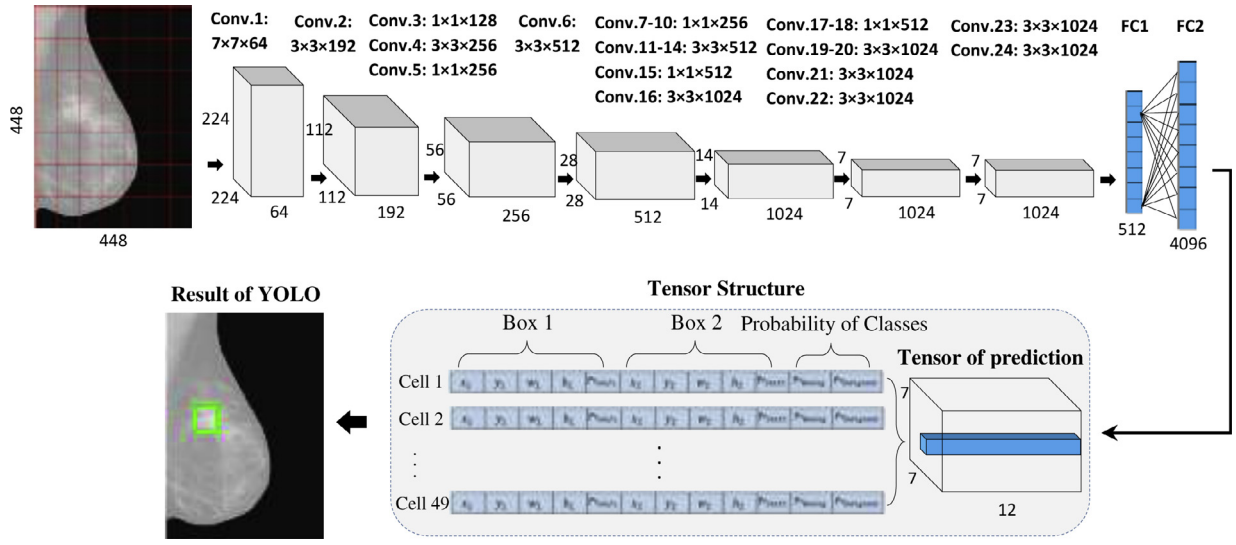

Fig. 2. The structures of proposed YOLO-based CAD system.

looks the image once and does not require a complex pipeline, it is extremely fast and its predictions are informed by global context in the data.

2.6. YOLO architecture

YOLO is a unified system that is able to detect the potential ROIs and directly predict their class probabilities from an entire whole image [15,25]. Our proposed YOLO-based CAD highlights two main issues of finding out the mass locations and classification their types of benign or malignant.

YOLO starts with dividing an input mammogram into $N \times N$ non-overlapped grid cells. Thus, each grid cell is responsible to detect the potential mass belonging to that cell. As successfully applied in [15], two bounding boxes with their confidence scores are utilized to represent each grid cell. Confidence is expressed as the probability of the existing mass multiplied with the percentage of the intersections over union (IOU) between the ground truth and the predicted boxes as follows:

$$\text{Confidence} = \text{Prob}(\text{mass}) \times \text{IOU}^{\text{ground truth}}_{\text{predicted}}. \quad (1)$$

Also, the detected mass is recognized as benign or malignant depending on the conditional class probability $\text{Prob}(\text{Class}_i|\,\text{mass})$ for the corresponding cell [15]. Then, the confidence score for each specific class is estimated as follows:

$$\begin{aligned}\text{Confidence score} &= \text{Prob}(\text{Class}_i|\,\text{mass}) \times \text{Confidence} \\ &= \text{Prob}(\text{Class}_i) \times \text{IOU}^{\text{ground truth}}_{\text{predicted}}, \end{aligned} \quad (2)$$

where confidence score interprets model confidence in order to represent the mass that is involved in the predicted box and also how accurate of that mass is. This confidence score becomes zero when the grid cell does not contain any objects. YOLO is trained utilizing the entire breast image with its ROIs' information. For training, we prepare the training data with the ROI position and size information: the information of training data contains the center position (x, y), width (w), height (h), and class label of the masses.

In this study, we use 24 convolutional layers with different kernel sizes, max-pooling layers with a size of 2×2, activation functions, and two fully-connected layers as inspired by [15,26]. The details of our proposed YOLO-based CAD structure is shown in Fig. 2. Deep based feature maps are extracted from each convolutional layer by applying different filter types (i.e., different kernels $\boldsymbol{W}$). The main role of using convolutional filters is to extract different features from the entire mammograms and then generates the feature maps. Convolution operation of $\boldsymbol{Y}_L^k$ that represents the kth feature map of layer L is computed as follows,

$$\boldsymbol{Y}_L^k = \phi\left(\boldsymbol{W}_L^k * \boldsymbol{Y}_{L-1}^k + b_L^k\right), \quad (3)$$

where $\phi(\bullet)$, '$*$', and b_L^k are the activation function, convolution operator, and bias for each feature map, respectively. In order to reduce the dimensionality of the features and select the proper features at each layer of the network, downsampling by max-pooling method is applied. Thus, only a maximum value from a 2×2 window is considered as an input for the next layer. In order to ensure that the features information is not lost, max-pooling filter size should not be large [6,15,27,28]. A pixel stride of two is utilized with all convolutional and pooling layers. The final aggregated deep features, which are produced by the convolutional and pooling layers, are passed to the fully-connected network (FC-NN). Linear leaky rectified activation function is used for all layers [15] and the Rectified Linear Unit (ReLU), $\phi(z) = \max(0,\ z)$, is only used for the final layer [6,29,30]. The leaky rectified activation function is defined to represent the linear transformation model of an input z as follows,

$$\phi(z) = \begin{cases} z, & \text{if } z > 0 \\ 0.1z, & \text{otherwise.} \end{cases} \quad (4)$$

Throughout the training phase, the training weights are updated through the fully-connected neural network layers utilizing training set images. According to [15] a batch size of 64 and learning rate of 0.001 are utilized to build our proposed CAD system. Finally, tensor of prediction (ToP) with size of $N \times N \times (5 \times M + C)$ is generated, where $N \times N$, M, and C are the number of grid cells, bounding boxes, and classes (i.e., benign and malignant), respectively. All of these parameters are selected as follows. Since we have two classes (i.e., benign and malignant), we set $C = 2$. Meanwhile, each grid cell becomes a unit which is responsible for detection and classification [15]. One can utilize different sizes, but we have chosen a size of 7×7 (i.e., $N = 7$) which gave the best performance compared to other sizes as investigated in [15]. To get the best predicted box among the inner and outer boundary of the object in the mammogram, we have selected $M = 2$. The bounding box with the highest confidence score was selected as the predicted box. Thus, the final output of the YOLO network represents a 3D matrix of ToP with size of $7 \times 7 \times 12$ as illustrated in Fig. 2.

Each grid cell of the entire mammogram is expressed by 12 elements in the tensor. The first five elements are corresponding to the predictions of the first bounding box, while the second five elements are for the second bounding box. For each box, these elements represent the prediction information of the mass locations which are *x, y, w, h*, and confidence probability. The last two elements (i.e., Pr_{Benign} and $Pr_{Malignant}$) in the ToP represent the confidence scores of the class probabilities for both benign and malignant cases, respectively. These class probabilities are considered for the highest confidence probability (i.e., the highest IOU with ground truth) among the bounding boxes. Thus, YOLO predicts only one bounding box for each grid cell which is responsible to detect the mass location and assign its appropriate class. Finally, among all of the potential predicted masses in each mammogram, YOLO only selects the boxes with confidence scores greater than a particular threshold. The Darknet framework is utilized for all training and testing processes [31].

2.7. Training and testing

In this paper, we have trained and tested our proposed CAD system using two different datasets (i.e., original and augmented datasets). All results of both detection and classification the breast abnormalities are obtained by training YOLO with the augmented dataset. In one exception, we compare the effect of data augmentation against the case of using the original dataset in Section 3.4. To avoid any bias in training and testing, we first optimized the parameters of the proposed YOLO-based CAD system using only the training dataset (i.e., 80% of the data). Then, the final system performance was evaluated using only the testing dataset (i.e., 20% of the data) [32]. It is shown that the concept of transfer learning is effective in training a deep net as in [14,15,33–35]. As this transfer learning was applied to DDSM in [6,14], we trained our YOLO-based CAD system with the pre-trained weights with a large computer vision ImageNet dataset [36]. Subsequently, it was fine-tuned (i.e., re-trained) with the training augmented mammograms.

To validate our results, we performed a *k*-fold cross validation ($k=5$) to ensure that every mammogram in our dataset gets to be in a test set exactly once and to minimize the bias error that may occur during the classification stage. The dataset is randomly divided into five subsets where each subset is formed by 10% benign and 10% malignant cases. One of the subsets (i.e., 20% of dataset) is utilized as a testing set while the other subsets (i.e., 80% of dataset) are considered together as a training set. This means we trained our YOLO-based CAD system five times to get the performance of the CAD system. For each *k*-fold, the computation time took almost four days to perform the training stage. However, the decoding (i.e., detection and classification) for each mammogram takes only less than three seconds. Thus, the proposed CAD system seems to be feasible and reliable to apply in the future for clinical applications. This wok was conducted on a PC Intel Core(TM) i5-3550 with 16 GB RAM, clock speed or frequency of CPU @ 3.30 GHz, and GPU of NVIDIA GeForce GTX 960. In addition, we utilized Python 2.7.6 and C++ as programming languages on operating system of Ubuntu 14.04. The results for both masses detection and classification are evaluated as an average of the 5-fold cross validation results.

2.8. Performance evaluation measures

In this study, we used objective measures to evaluate the performance of our YOLO-based CAD system. Fig. 3 shows our evaluation logic during the testing phase of detecting the mass location in the mammogram and classifying its type into benign or malignant. According to this logic, if the confidence probability scores of the detected boxes are less than a particular threshold, the corresponding predicted ROIs or masses are considered as undetectable cases as shown in Fig. 3. This means that these cases are excluded during the next classification and detection stages. In contrast, the predicted masses which have confidence probability scores equal or greater than this threshold are considered for the next classification and detection assessment processes. For detection assessment, the mass location is properly detected if and only if IOU equals or exceeds 50% comparing with its ground truth. In the case of false mass detection, as long as the condition of threshold probability is satisfied, the final decision to distinguish the type of these masses is achieved utilizing the capability of the classifier.

To quantitatively present the capability of the proposed technique, confusion matrix is utilized to show how the proposed CAD is able to distinguish between benign and malignant classes [17]. Meanwhile, the curve of receiver operator characteristic (ROC) with its area under curve (AUC) are used for classification evaluation purposes as well. AUC of ROC shows the performance of the classifiers where AUC value close to 1.0 represents the highly accurate diagnostic rate while AUC value close to 0.5 indicates the unreliable performance. ROC curve definition is as follows [6,17],

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP}+\text{FN}}, \tag{5}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN}+\text{FP}}, \tag{6}$$

where, TP and FN denote the true positive and false negative cases, respectively. TN and FP contain the true negative and false positive cases, respectively. The overall classification accuracy of the system is defined as,

$$\text{Overall accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{FN}+\text{TN}+\text{FP}}. \tag{7}$$

In addition, the accuracy of the detected masses in mammograms compared to the ground truth is quantitatively evaluated using free response operating characteristic (FROC) curve [14,37,38]. FROC curve is a function which represents the true positive detection rate versus the false positive rate per image.

3. Results

3.1. Class probability threshold

As aforementioned, the confidence score for each grid cell represents the class probability of the highest confidence among the two bounding boxes. Therefore, the proposed YOLO-based CAD system generates many potential ROIs for each testing mammogram. We attempted to find the appropriate threshold that avoids the unwanted ROIs with too small class probability. For instance, all potential ROIs are shown on the mammogram in the case of zero threshold as shown in Fig. 4(a). In order to maintain the testing data (i.e., less number of undetectable data as explained in Fig. 3), threshold should not be large. Thus, we investigated the appropriate threshold that achieves the minimum number of undetectable data. Fig. 4(b) and (c) show the potential ROIs with class probability threshold of 0.01 and 0.2, respectively. Boxes with higher confidence are illustrated with thicker border. In this study, we experimentally determined a probability threshold to be 0.2. As it is explained above, this threshold controls the number of potential ROIs through each testing mammogram. With the probability threshold of 0.2, at least one potential ROI was provided while ignoring all undesirable ROIs with too small class probability. A proper threshold must provide enough detected ROIs for further classification.

3.2. Mass detection via the YOLO-based CAD

The results of the proposed CAD system on the ability of detecting the location of the benign and malignant masses are shown in

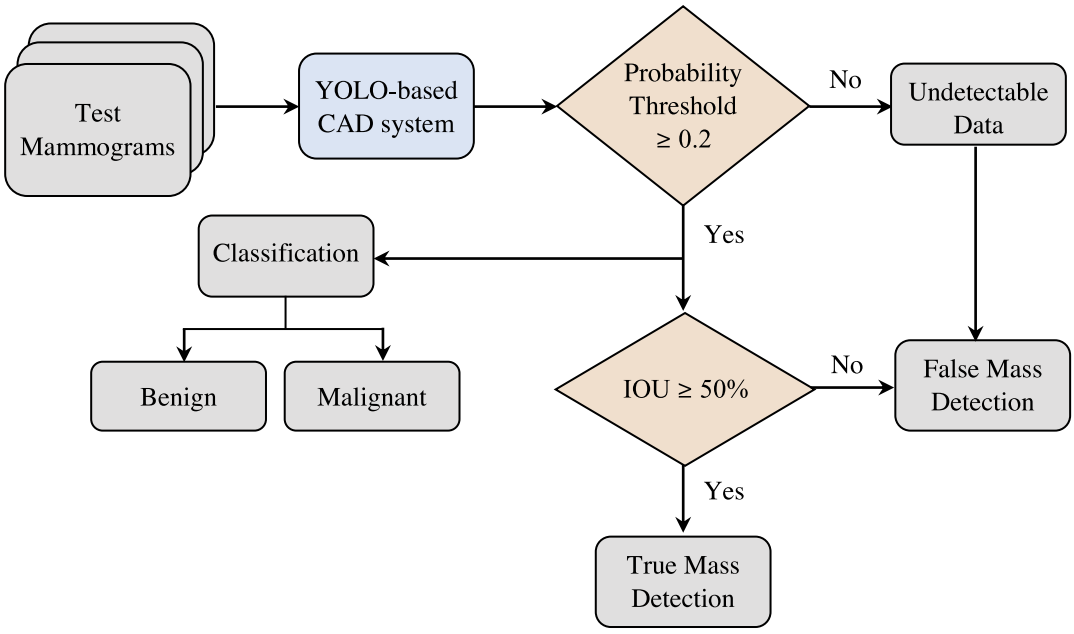

Fig. 3. Present study evaluation approach for both mass detection and classification.

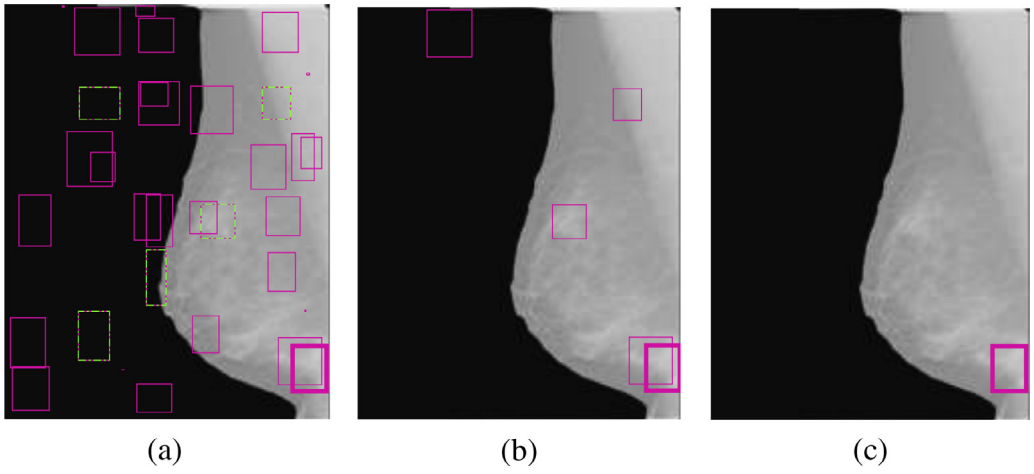

Fig. 4. Potential ROIs with a threshold probability of (a) zero, (b) 0.01, and (c) 0.2. The ROI with the highest confidence has the thickest border.

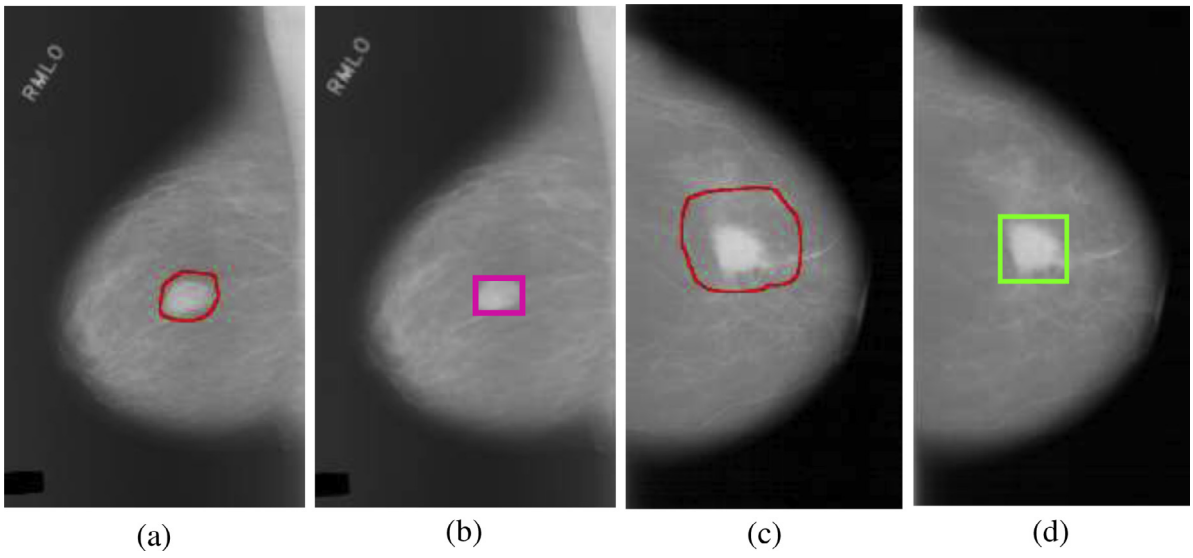

Fig. 5. Mass detection. (a) and (b) show the ground-truth mass and detected from new proposed method for a benign case, while (c) and (d) for a malignant case.

Table 1
5-fold cross validation performance of the mass detection via the proposed YOLO-based CAD system.

Fold test	Benign		Malignant		Total	
	True	False	True	False	True	False
1st fold	**240**	**0**	**240**	**0**	**480**	**0**
	100%	0.0%	100%	0.0%	100%	0.0%
2nd fold	**237**	**3**	**239**	**1**	**476**	**4**
	98.75%	1.25%	99.58%	0.42%	99.17%	0.83%
3rd fold	**238**	**2**	**240**	**0**	**478**	**2**
	99.17%	0.83%	100%	0.0%	99.58%	0.42%
4th fold	**239**	**1**	**240**	**0**	**479**	**1**
	99.58%	0.42%	100%	0.0%	99.79%	0.21%
5th fold	**240**	**0**	**240**	**0**	**480**	**0**
	100%	0.0%	100%	0.0%	100%	0.0%
Average (%)	**99.50**	**0.50**	**99.92**	**0.08**	**99.71**	**0.29**

Fig. 5. Fig. 5(a) and (c) show the ground truth and Fig. 5(b) and (d) show the masses detected by our YOLO-based CAD. The abnormalities (i.e., masses) detection performance throughout the 5-fold test of benign and malignant is reported in Table 1. At each fold test, a test dataset contains 480 mammograms which are equally divided into benign and malignant cases. The results show the robustness of our YOLO-based CAD on detecting the mass position in mammograms with an overall accuracy of 99.7%. As mentioned above, YOLO generates confidence probability for each potential ROI (indicated as a box) that represents the mass position. In Table 1, false detection cases represent those have the detected boxes with confidence probability less than 0.2 or IOU less than 0.5. These detection results are achieved by comparing the IOU of each detected boxes with the ground truth. Fig. 7(a) shows the detection performance of the proposed YOLO-based CAD system through 2nd fold test in term of FROC curve. The potential detected boxes are considered as successfully detected if the overlap equals or exceeds 50% comparing with their corresponding ground truth masses. The proposed method in Fig. 7(a) produces a true positive rate of 99.17% at false positive per image equals to 0.22.

3.3. Mass classification via the YOLO-based CAD

For classification, an overall accuracy is computed using all test data. Only the undetectable mammograms in the detection stage throughout each *k*-fold subset are excluded from the evaluation of masses prediction. Fig. 5 shows some representative results of the YOLO-based CAD system in terms of detection and classification for two cases: benign and malignant. The breast regions that are not detected are addressed as normal tissue. Confusion matrices for each *k*-fold subset of the YOLO-based CAD system are presented in Table 2. It is clearly shown that the benign cases are correctly classified with 100% in all k-fold subsets, while the malignant cases are in between 92.5% and 95.8%. The false positive in the 1st fold subset represents the 18 malignant cases that are incorrectly classified and negatively affect the specificity. It is clear from Table 2 that the results of AUCs and accuracies of all *k*-fold are similar to each other. This indicates the efficiency and feasibility of our proposed CAD system on diagnosis of the breast masses. We summarize the performance of the proposed CAD system as an average of the 5-fold cross validation results in terms of sensitivity, specificity, AUC, and overall detection and classification accuracies as follows. The YOLO-based CAD system performed an overall detection and classification accuracies of 99.70% and 97.00%, respectively. The sensitivity and specificity values reflect the statistical measures of true positive and true negative rates, respectively. Our CAD system achieved sensitivity for benign cases with 100% and specificity for malignant cases with 94%.

The masses that exist over the pectoral muscle or are surrounded by dense tissue in the mammograms are generally known as well-known challenges for mass detection and classification. These challenges are generally due to the variation in shape, texture, and size of the different masses [4]. Fig. 6(a) and (c) show such cases and their results of our proposed YOLO-based CAD in Fig. 6(b) and (d). Fig. 6(b) and (d) illustrate the capability of our proposed CAD to detect and predict the mass in these two challenging cases. Our proposed CAD system seems to overcome these challenges.

3.4. The effect of the size of data sets

In this section, we present the effect of different training dataset sizes (i.e., original vs. augmented datasets) that are utilized in training of our YOLO-based CAD system. The original dataset contains only 600 mammograms. However, the augmented dataset involves all original and augmented mammograms which are 2400 cases. In general, the performance of deep learning techniques improves as the size of the training dataset increases. Fig. 7(b) illustrates the effect of enlarged dataset on the proposed CAD system. It shows the improvement in term of ROC curves in the case of the augmented dataset against the original ones with AUCs of 96.45% and 87.74%, respectively. Significantly, the augmented data also affects the specificity with an improvement of 16% as reported in Table 3. Due to this, the overall classification accuracy is increased from 85.52% to 97%. In addition, YOLO-based CAD system presents stability with slightly improvement rate regarding the performance of mass detection. Note that the presented results in all sections are based on the training of the augmented dataset.

4. Discussion

In this study, we have developed a deep learning YOLO-based CAD system which detects the locations of potential masses on mammograms and classifies them into benign or malignant simultaneously. The recent deep learning CAD systems only addressed the diagnosis task of the extracted patches from mammograms [6,7,10,11,13,23]. In contrast, the proposed YOLO-based CAD system could handle both detection and classification at the same time using whole breast image. Figs. 5 and 6 show the capability of the proposed CAD system to detect the potential breast masses and produce the proper diagnosis for each mammogram (i.e., benign or malignant). Furthermore, we observe that YOLO-based CAD system overcomes two important challenges faced CAD approach in the clinical mammographic field. First, it could reveal the breast masses which are existed over the pectoral muscle as shown in Fig. 6(b). Second, the proposed methodology successfully identified breast masses in the dense tissues as shown in Fig. 6(d). In fact, both of these challenges are due to the high intensities (i.e., more bright) among the pectoral muscle and dense tissue regions compared to normal breast tissue. The results of Tables 1 and

Table 2
Confusion matrices and performance of our YOLO-based CAD system throughout 5-fold cross validation.

Fold test	Actual classes	Predicted classes		Sensitivity (%)	Specificity (%)	AUC (%)	Accuracy (%)
		Benign	Malignant				
1st fold	**Benign**	**240**	**0**	100	92.5	95.73	96.25
		100%	0.0%				
	Malignant	**18**	**222**				
		7.5%	92.5%				
2nd fold	**Benign**	**239**	**0**	100	95.82	97.16	97.91
		100%	0.0%				
	Malignant	**10**	**229**				
		4.2%	95.8%				
3rd fold	**Benign**	**240**	**0**	100	94.58	96.83	97.29
		100%	0.0%				
	Malignant	**13**	**227**				
		5.4%	94.6%				
4th fold	**Benign**	**239**	**0**	100	94.17	95.85	97.08
		100%	0.0%				
	Malignant	**14**	**226**				
		5.8%	94.2%				
5th fold	**Benign**	**240**	**0**	100	92.92	96.66	96.46
		100%	0.0%				
	Malignant	**17**	**223**				
		7.1%	92.9%				
Average	**Benign**	**100%**	**0.0%**	**100**	**94.00**	**96.45**	**97.00**
	Malignant	**6.00%**	**94.00%**				

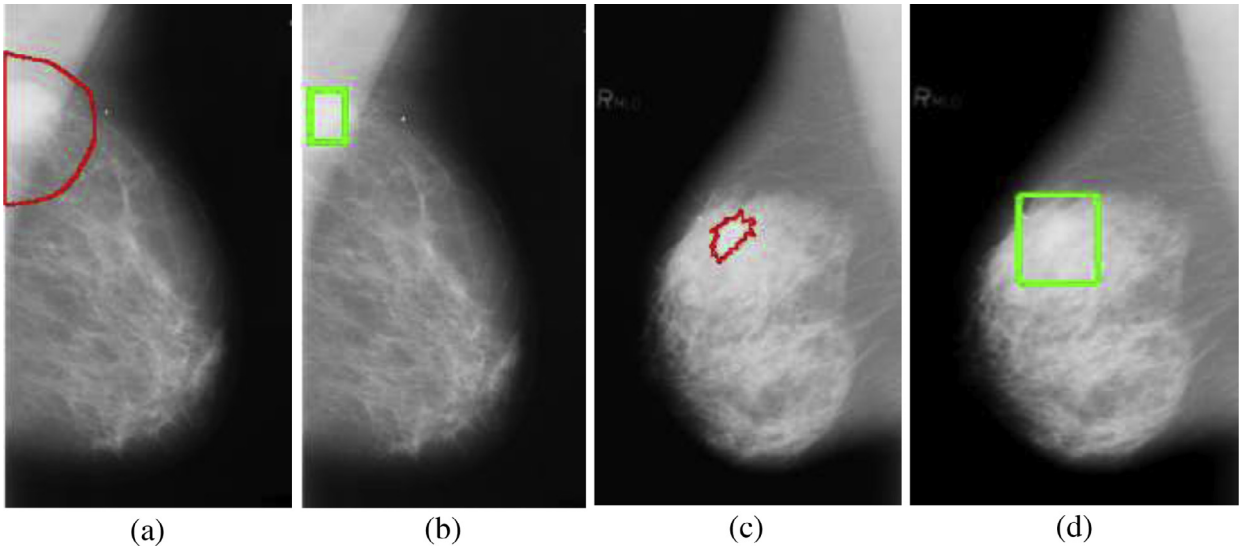

Fig. 6. Mass detection and classification. (a) and (c) show the ground-truth mass over the pectoral muscle and the detected by new proposed method, respectively. (b) and (d) present the ground-truth mass surrounding by dense tissue and the detected by new proposed CAD.

Table 3
Comparison of the performance (%) of the effect of the dataset size via present proposed YOLO-based CAD system.

Index	YOLO–CAD with original dataset	YOLO–CAD with augmented dataset
Sensitivity	93.20	100.00
Specificity	78.00	94.00
AUC	87.74	96.45
Classification accuracy	85.52	97.00
Detection accuracy	96.33	99.70

2 show the analysis of both detection and classification of the breast masses throughout 5-fold cross validation.

It is previously shown that training with augmented data improves the performance of breast masses detection and classification [6,14,18,19]. The overall accuracy performance increased from 85.5% with the 600 original mammograms to 97% with the 2400 augmented mammograms as shown in Fig. 7(b) and Table 3. These results demonstrate that the YOLO-based CAD system is effective to achieve high accuracy in both detection and classification of the abnormalities at the same time.

In order to show how robust the proposed YOLO-based CAD system is, we compare the results with the latest studies employing DBN and CNN. Comparison with the conventional classifiers that do not utilize deep learning is also provided to present the efficiency of the deep learning algorithms. In the previous work [23] that utilized same kind of DDSM mammograms, a DBN-based CAD system was applied and compared its outcomes against the conventional Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and Neural Network (NN) classifiers. Statistical handcrafted features are excerpted from the extracted masses. That study investigated the effect of features dimensionality reduction through different feature selection methods. Table 4 shows the overall classification accuracies by utilizing the Sequential Floating Forward (SFFS) algorithm. In the case of DBN tech-

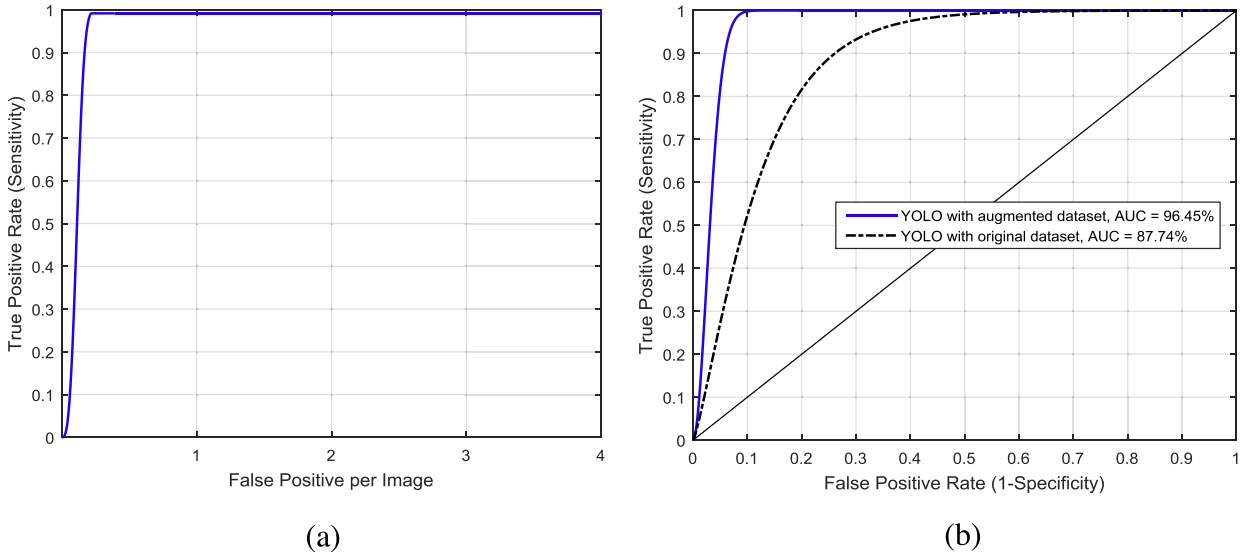

Fig. 7. (a) FROC curve performance of breast mass detection. (b) ROC curves of the proposed YOLO-based CAD system with the augmented dataset against the original dataset.

Table 4
Classification performance of present proposed YOLO-based CAD system against conventional classifiers, DBN, and CNN.

Reference	Method	Database (No. images)	Prediction classes	Mass detection accuracy (%)	Classification accuracy (%)
Al-antari et al. [23]	LDA	DDSM (168)	Normal / Benign / Malignant	86.00	78.57
	QDA				76.19
	NN				84.52
	DBN				90.48
Jiao et al. [6]	CNN	DDSM (2400)	Benign / Malignant	X	96.70
Present study	**YOLO**	**DDSM (2400)**	**Benign / Malignant**	**99.7**	**97.00**

nique, all the extracted features are utilized, without the need of features reduction, to train and test the CAD system. It is clearly shown that how the DBN overcomes the conventional classifiers with overall accuracy of 90.48%. Also, a comparison study between random forest (RF) classifier against CNN was investigated in [18]. Different kinds of features set are manually extracted from the mass patches to train RF classifier. Each features set are individually trained and then they applied them separately to the test set. Their results showed that the features group of candidate detector, contrast, texture, geometry, location, context, and patient information got AUCs of 85.8%, 78.7%, 71.8%, 75.3%, 68.6%, 81.6%, and 65.1%, respectively. While all the feature sets together obtained AUC of 90.6% against 92.0% in the case of CNN. In addition, a framework for CAD system utilizing CNN technique was presented in [6]. As different of our proposed CAD system that utilized the whole mammograms for the convolutional layers, they only used the ROIs of the cropped masses. Combination of the middle level and high level features are utilized to train and test the CAD system based CNN. The performance of CNN shows it capability to classify the masses into benign or malignant with overall accuracy of 96.7%. Actually, this classification results are highly comparable with ours. In contrast, only our proposed YOLO-based CAD system can detect the masses in mammograms besides predicting their types compared with the conventional CNN. Our proposed technique should be feasible as a CAD system capable of detection and classification the abnormalities of the breast images.

Finally, we present a comparison of the effect of utilizing augmented data instead of original ones.

In [18], the mass dataset are augmented utilizing three transformation types: rotation, translation, and scaling. Normal and malignant cases are classified by applying CNN to the mass patches with a size of 250×250. Their AUC results of the CNN without the augmented dataset achieved 87.5%, while it reached to 92.9% with the augmented dataset. This improvement rate is comparable with our AUC results from 87.74% to 96.45%.

5. Conclusion

In this paper, we present YOLO-based CAD system for breast mass detection and cancer classification. The proposed CAD system incorporates a ROI-based CNN approach which utilizes the convolutional layers followed by fully connected neural networks to detect the proper location of the mass and to distinguish the tumor types: benign or malignant. Our results provide feasible and promising results in term of detecting the location of benign and malignant masses and recognize their proper classes as well. Furthermore, the YOLO-based CAD system detects the masses existing over the pectoral muscle or surrounding by the dense tissue in the mammograms which are considered as most challenging cases of breast cancer CAD. The next step of the presented CAD system is to be tested in practice for its real validity.

Acknowledgments

This work was supported by the Center for Integrated Smart Sensors funded by the Ministry of Science, ICT & Future Planning as a Global Frontier Project (CISS- 2011–0031863). This work was also supported by International Collaborative Research and Development Programme (funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea) (N0002252).

References

[1] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2016, CA Cancer J. Clin. 66 (1) (2016) 7–30.

[2] J. Virmani, N. Dey, V. Kumar, in: PCA-PNN and PCA-SVM Based CAD Systems For Breast Density Classification, Springer International Publishing, Warsaw, Poland, 2016, pp. 159–180.
[3] C. Muramatsu, T. Hara, T. Endo, H. Fujita, Breast mass classification on mammograms using radial local ternary patterns, Comput. Biol. Med. 72 (1) (2016) 43–53.
[4] H. Li, X. Meng, T. Wang, Y. Tang, Y. Yin, Breast masses in mammography classification with local contour features, BioMed. Eng. OnLine 16 (1) (2017) 44–54.
[5] S.A. Taghanaki, J. Kawahara, B. Miles, G. Hamarneh, Pareto-optimal multi-objective dimensionality reduction deep auto-encoder for mammography classification, Comput. Methods Progr. Biomed. 145 (2017) 85–93.
[6] Z. Jiao, X. Gao, Y. Wang, J. Li, A deep feature based framework for breast masses classification, Neurocomputing 197 (C) (2016) 221–231.
[7] J. Arevalo, F.A. González, R. Ramos-Pollán, J.L. Oliveira, M.A.G. Lopezd, Representation learning for mammography mass lesion classification with convolutional neural networks, Comput. Methods Progr. Biomed. 127 (2016) 248–257.
[8] N. Dhungel, G. Carneiro, A.P. Bradley, Automated mass detection in mammograms using cascaded deep learning and random forests, 2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA), 2015.
[9] N. Dhungel, G. Carneiro, A.P. Bradley, A deep learning approach for the analysis of masses in mammograms with minimal user intervention, Med. Image Anal. 37 (2017) 114–128.
[10] T. Kooi, A. Gubern-Merida, J.-J. Mordang, R. Mann, R. Pijnappel, K. Schuur, A. d. Heeten, N. Karssemeije, A comparison between a deep convolutional neural network and radiologists for classifying regions of interest in mammography, International Workshop on Digital Mammography, 2016.
[11] M.A. Al-antari, M.A. Al-masni, S.U. Park, J.H. Park, Y.M. Kadah, S.M. Han, T.-S. Kim, Automatic computer-aided diagnosis of breast cancer in digital mammograms via deep belief network, Global Conference on Engineering and Applied Science (GCEAS), 2016.
[12] A.-B. Ayelet, L. Karlinsky, S. Alpert, S. Hasoul, R. Ben-Ari, E. Barkan, A region based convolutional network for tumor detection and classification in breast mammography, in: International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, Athens, Greece, Springer International Publishing, 2016, pp. 197–205.
[13] Y. Qiu, S. Yan, R.R. Gundreddy, Y. Wang, S. Cheng, H. Liu, B. Zheng, A new approach to develop computer-aided diagnosis scheme of breast mass classification using deep learning technology, J. X-Ray Sci. Technol. 25 (5) (2017) 751–763.
[14] G. Carneiro, J. Nascimento, A.P. Bradley, Automated analysis of unregistered multi-view mammograms with deep learning, IEEE Trans. Med. Imag. 36 (11) (2017) 2355–2365.
[15] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: unified, real–time object detection, IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[16] M. Heath, K. Bowyer, D. Kopans, R. Moore, W.P. Kegelmeyer, The digital database for screening mammography, 5th international Workshop on Digital Mammography, 2000.
[17] M. Dong, X. Lu, Y. Ma, Y. Guo, Y. Ma, K. Wang, An efficient approach for automated mass segmentation and classification in mammograms, J. Digit. Imag. 28 (5) (2015) 613–625.
[18] T. Kooi, G. Litjens, B. v. Ginneken, A. Gubern-Mérida, C.I. Sánchez, R. Mann, A. d. Heeten, N. Karssemeijer, Large scale deep learning for computer aided detection of mammographic lesions, Med. Image Anal. 35 (2017) 303–312.
[19] H.R. Roth, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, R.M. Summers, Improving computer-aided detection using convolutional neural networks and random view aggregation, IEEE Trans. Med. Imag. 35 (5) (2016) 1170–1181.
[20] T. Wu, R.H. Moore and D.B. Kopans, "Multi-threshold peripheral equalization method and apparatus for digital mammography and breast tomosynthesis," *U.S. Patent 7,764,820*, 2010.
[21] M.A. Al-antari, M.A. Al-masni, Y.M. Kadah, Hybrid model of computer-aided breast cancer diagnosis from digital mammograms, J. Scientific Eng. 04 (02) (2017) 114–126.
[22] M. Kallenberg, N. Karssemeijer, Comparison of tilt correction methods in full field digital mammograms, in: Digital Mammography/IWDM, Springer, Catalonia, Spain, 2010, pp. 191–196.
[23] M.A. Al-antari, M.A. Al-masni, S.U. Park, J.H. Park, M.K. Metwally, Y.M. Kadah, S.M. Han, T.-S. Kim, An automatic computer-aided diagnosis system for breast cancer in digital mammograms via deep belief network, J. Med. Biol. Eng. (2017). https://doi.org/10.1007/s40846-017-0321-6.
[24] A. Coates, A. Ng, H. Lee, An analysis of single-layer networks in unsupervised feature learning, Fourteenth International Conference on Artificial Intelligence and Statistics, 2011.
[25] M.A. Al-Masni, A.-A.A. Mugahed, J.M. Park, G. Gi, T.Y. Kim, P. Rivera, E. Valarezo, S.-M. Han, T.-S. Kim, Detection and classification of the breast abnormalities in digital mammograms via regional convolutional neural network, Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE, 2017.
[26] M. Lin, Q. Chen and S. Yan, "Network in network," in *arXiv:*1312.4400, 2013.
[27] C. Cernazanu-Glavan, S. Holban, Segmentation of bone structure in X-ray images using convolutional neural network, Adv. Electr. Comput. Eng. 13 (1) (2013) 87–94.
[28] D. Scherer, A. Müller, S. Behnke, Evaluation of pooling operations in convolutional architectures for object recognition, Artificial Neural Networks–ICANN, 2010.
[29] Q. Guo, F. Wang, J. Lei, D. Tu, G. Li, Convolutional feature learning and Hybrid CNN-HMM for scene number recognition, Neurocomputing 184 (2016) 78–90.
[30] J.H. Park, S.U. Park, M. Zia Uddin, M. Al-antari, M. Al-masni, T.-S. Kim, A single depth sensor based human activity recognition via convolutional neural network, 4th World Conference on Applied Sciences, Engineering & Technology, 2016.
[31] J. Redmon, "Open source neural networks in c," http://pjreddie.com/darknet/, 2013-2016.
[32] E. Szymańska, E. Saccenti, A.K. Smilde, J.A. Westerhuis, Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies, Metabolomics 8 (1) (2012) 3–16.
[33] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, H. Greenspan, Chest pathology identification using deep feature selection with non-medical training, Comput. Methods Biomech. Biomed. Eng. 13 (2016) 1–5.
[34] R.K. Samala, H. Chan, L. Hadjiiski, M.A. Helvie, J. Wei, K. Cha, Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography, Med. Phys. 43 (12) (2016) 6654–6666.
[35] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How transferable are features in deep neural networks? Advances in Neural Information Processing Systems, 2014.
[36] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, ImageNet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (3) (2015) 211–252.
[37] C.E. Metz, Receiver operating characteristic analysis: a tool for the quantitative evaluation of observer performance and imaging systems, J. Am. College. Radiol. 3 (6) (2006) 413–422.
[38] D.P. Chakraborty, K.S. Berbaum, Observer studies involving detection and localization: modeling, analysis, and validation, Med. Phys. 31 (8) (2004) 2313–2330.